1

One-time programmable memory device

5

The invention relates to a one-time programmable (OTP) memory device and to a method for programming a one-time programmable memory device.

An OTP memory constitutes a non-volatile storage element which preserves an information even if a power supply is turned off. Conventional OTP

10 memories have a structure which differs from the structures used in conventional CMOS (complementary metal-oxide semiconductor) technology. Thus, they require modifications or expensive extra processing steps in wafer fabrication when they are to be integrated with CMOS circuitry.

15

In document US 5,943,264, an approach for realizing an OTP memory device in standard CMOS technology is presented. Here, an NMOS (N-channel metal-oxide semiconductor) transistor is connected in series with a PN junction, which PN junction functions as memory. The memory is programmed by fusing the PN junction.

20 To this end, the transistor is brought into a so-called snap-back mode, which is achieved when the drain-source voltage reaches a specific breakdown voltage. In the snap-back mode, the current gain of the transistor is very high and its series resistance is very low. Thereby, the transistor is able to provide a high current suited to fuse the PN junction, even in case the transistor is rather small. This approach has the disadvantage, however,

25 that the transistor itself might be damaged in the snap-back mode. It is therefore proposed that additional programming current limitation means are provided to prevent the destruction of the transistor.

30 It is an object of the invention to provide a particularly simple and

reliable alternative for programming a one-time programmable memory.

This object is reached according to the invention with an OTP memory device comprising a MOS selection transistor and a MOS memory transistor connected in series between a voltage supply line and ground. The device further comprises

5   programming means for applying predetermined voltages to the gate of the selection transistor, to the gate of the memory transistor and to the voltage supply line, which applied voltages force the memory transistor into a snap-back mode resulting in a current thermally damaging the drain junction of the memory transistor.

The object of the invention is equally reached with a corresponding

10   method for programming an OTP memory device comprising a MOS selection transistor and a MOS memory transistor connected in series between a voltage supply line and ground.

The invention proceeds from the idea that if a MOS transistor is employed as memory, this memory transistor can be programmed by being brought

15   itself into a snap-back mode, in which a drain junction of the transistor is thermally damaged. Thereby, the OTP memory device can be programmed without the necessity to apply voltages exceeding the snap-back holding voltage of the process technology.

It is an advantage of the invention that the employed memory element can be programmed without stressing a selection transistor providing the current for

20   programming the memory transistor.

The invention provides a low-cost option for OTP memory applications which are based on CMOS technology. No changes of the CMOS fabrication process are required.

Preferred embodiments of the invention become apparent from the

25   dependent claims.

Advantageously, the memory transistor is forced into the snap-back mode by ramping down the voltage which is first applied to the gate of the memory transistor. Thereby,

the effectiveness of the drain fusing mechanism is increased, which

30   limits the required size of the selection transistor.

The employed memory transistor should be rather small in order to keep

the required programming current low, while the selection transistor has to be large enough to provide the required programming current to the memory transistor. The required programming current can be about 10mA, but varies with the used technology. Due to the required size of the selection transistor, the invention is particularly suited

5   for applications which only require a small number of bits to be stored.

The memory device preferably meets a variety of requirements which ensure a sufficient reliability of the memory device. Such requirements comprise ESD (electrostatic discharge) robustness, latch-up robustness, minimal decoder and selection transistor degradation, data retention of non-programmed cells and data retention of

10   programmed cells.

ESD stress, which usually occurs during the handling of a device when no power supply is connected to it, might fuse memory transistor drains in case no specific protection is provided. In a preferred embodiment of the invention, therefore an RC unit is implemented in the memory device. The RC time constant of this unit

15   ensures that a voltage applied to the gate of the selection transistor and a voltage applied to the gate of the memory transistor are low during a time period exceeding the usual duration of an ESD event when the memory device is powered up. That is, the gates of these transistors will remain grounded with a timer period of e.g. 10 ms after powering the memory device up, which ensures that a high voltage occurring at a programming

20   pad of the programming means during an ESD event will not reach the memory transistor. Thus, the proposed embodiment of the invention provides an effective protection against ESD stress.

A latch-up, in contrast, may occur during normal operation when the device is powered up. More specifically, there is a risk that an external overvoltage

25   pulse will activate the programming mode of the OTP memory device. Therefore, a mechanism should be provided which actively switches the programming current path off during normal operation. In a preferred embodiment of the invention, such a mechanism is given by a sufficiently complex setup procedure for activating the programming mode, which ensures with a sufficiently high probability that a single

30   external overvoltage pulse will not activate the programming mode.

In an equally preferred embodiment of the invention, a programming

voltage level above the normal operation voltage level is used, in order to keep the cell size reasonably small. High voltage levels may result in a degradation of the memory device due to an increased heating of the carrier. Hot carrier effects occur e.g. in strong pinch-off conditions in a transistor, which in turn occur with a high drain voltage and a

5   moderate to low gate voltage at the transistor. It has thus to be ensured that no intolerable degradation of the memory circuitry occurs due to the proposed high programming voltage. Hot carrier conditions resulting in a degradation, however, are avoided effectively with the proposed programming cycle comprising a ramping down of the voltage applied to the gate of the memory transistor.

10   The memory transistor should moreover be designed such that it can withstand the normal supply voltage, e.g. the voltage applied for a memory readout, without being damaged, in order to ensure data retention of a non-programmed memory transistor.

The programming of the memory transistor results in a leakage in the

15   transistor, which can be detected in the form of a leakage current in a subsequent readout. It should be ensured in the readout that a detected leakage is indeed caused by a hard drain fusing which results in a damage connecting the drain with the source diffusion, not only by an ESD-induced soft drain junction damage which may reduce over time. As the leakage currents resulting

20   from hard drain junction fusing are in the mA range, while those resulting from soft damage are in the µA range, they can easily be distinguished by setting a proper current detection limit for the readout.

The invention is particularly suited to be employed in any product which is designed in a CMOS technology and which requires a low-bit memory, for example

25   in watch ICs using an OTP memory for an adjustment of some oscillator frequency, in display driver chips using an OTP memory for an adjustment of voltage levels, etc.

Other objects and features of the present invention will become apparent from the following detailed description considered in conjunction with the accompanying drawings, wherein

30

Fig. 1   schematically illustrates a memory cell of a memory device according to the invention; and

Fig. 2   is a diagram illustrating the snap-back mode of an NMOS transistor.

5

Figure 1 shows an OTP memory cell forming part of a CMOS memory device according to the invention. The memory cell comprises a selection transistor T1 and a memory transistor T2. Both transistors T1, T2 are NMOS transistors. The drain of

10   the selection transistor T1 is connected to a bit line BL and the gate of the selection transistor T1 is connected to a voltage supply (not shown) providing a voltage Vsel. The source of the selection transistor T1 is connected to the drain of the memory transistor T2, while the source of the memory transistor T2 is connected to ground Gnd. The gate of the memory transistor T1 is connected to a voltage supply (not shown) providing a

15   voltage Vctrl.

In a non-programmed state of the memory cell, the drain junction of the memory transistor T2 is intact, while in a programmed state of the memory cell, the drain junction of the memory transistor T2 is thermally damaged.

For programming a memory cell, the voltages Vsel and Vctrl applied to

20   the gate of the selection transistor T1 and to the gate of the memory transistor T2, respectively, are set by programming means (not shown) to a voltage level predetermined for programming. The transistors T1, T2 are thus switched on. Further, a predetermined programming voltage Vprog is applied by the programming means to the bit line BL. The programming voltage Vprog is set to a level exceeding the allowed

25   maximum operating voltage of the employed CMOS technology, but below the snap-back voltage of the employed CMOS technology.

As a result of the applied voltages Vsel, Vctrl and Vprog, a current flows through the memory cell, i.e. from bit line BL via selection transistor T1 and memory transistor T2 to ground Gnd. The transistors T1, T2 are dimensioned such that the major

30   part of the voltage Vprog applied to the bit line BL is applied to the memory transistor T2. The selection transistor T1 is not required to be strong enough to provide a current

to the memory transistor T2 which is sufficient to thermally damage the drain junction of the memory transistor T2 at this point of time.

In order to ensure that the drain junction of the memory transistor T2 is thermally damaged, the voltage Vctrl applied to the gate of the memory transistor T2 is
5   now ramped down by the programming means. This forces the memory transistor T2 into a strong pinch-off and eventually into a snap-back mode, even though the drain-source voltage is below the snap-back voltage.

For illustration, figure 2 shows the drain current I_D over the drain-source voltage V_D of an NMOS transistor. A first curve 1 indicates the behavior of the
10 transistor for the case that the gate of the transistor is grounded. It can be seen that the transistor switches back, or snaps back, at the voltage Vsb and goes into a high-current mode at a lower snap-back holding voltage Vsbh. When the gate of the transistor is turned on, the transistor passes on directly to the high-current mode at snap-back holding voltage Vsbh without having to reach first the snap-back voltage Vsb. This is
15 indicated in figure 2 with a second curve 2. Thereby, the power dissipation is concentrated even further into the drain junction of the transistor, which provides an effective drain junction fusing mechanism.

Ramping down the voltage Vctrl which is applied to the gate of the memory transistor T2 further increases the effectiveness of the drain fusing mechanism.
20   A subsequent readout of the memory transistor T2 is performed by readout means (not shown) sensing a leakage current in the memory cell.

To this end, the readout means apply an operating voltage Vop to the bit line BL. Moreover, the voltage Vsel applied to the selection transistor T1 is set high, while the voltage Vctrl applied to the gate of the memory transistor T2 is set low. That
25 is, the selection transistor T1 is turned on with a gate-voltage predetermined for the readout, while the memory transistor T2 is not turned on.

In case the memory cell is not programmed, basically no current will flow through the cell, since the memory transistor T2 is intact and turned off. If the cell is programmed, in contrast, a current should flow through the cell even though the
30 memory transistor T2 is turned off, since the fused drain junction of the memory transistor T2 allows a leakage current to pass.

The amount of a current flow detected by a sensing circuitry of the reading means can thus be used as a criterion for determining whether the cell has to be considered to be programmed or not. A cell with a detected current lying below a predetermined detection level is considered not to be programmed, whereas a cell with a

5   detected current lying above a predetermined detection level is considered to be programmed.

Experiments have shown that drain fusing occurs without an additional gate oxide breakdown with a certainty level of 90%. But even in case the gate oxide of a programmed memory transistor has broken down, this cell will still allow a leakage

10 current and be considered as programmed. Therefore, the proposed memory is insensitive to gate oxide breakdown in a programmed cell.

A plurality of memory cells corresponding to the memory cell described with reference to figure 1 can be combined in a memory cell array. The memory cell area is dominated by the size of the selection transistor and its guard ring. Such a guard

15 ring is required to assure a sufficient ESD robustness of the memory cell. The space requirement for the selection transistor can be assumed to be about 20 μm x 8.2 μm = 164 μm^2, and the space requirement for the memory transistor can be assumed to be about 8 μm x 8 μm = 64 μm^2, resulting in an estimated cell size of about 220 μm^2. A 32-bit memory cell array would thus require an area of 7040 μm^2 or e.g. 70 μm x

20 100 μm. To this area, the area required for a multiplexer circuitry taking care of programming and reading out the different cells has to be added.

It is to be noted that the presented embodiment constitutes only a selected embodiment of the invention which can be varied in many ways. In particular, suitable protection mechanisms protecting a memory cell from an undesired programming can

25 provided.